

CS3245

Information Retrieval

Lecture 0: Course Organization



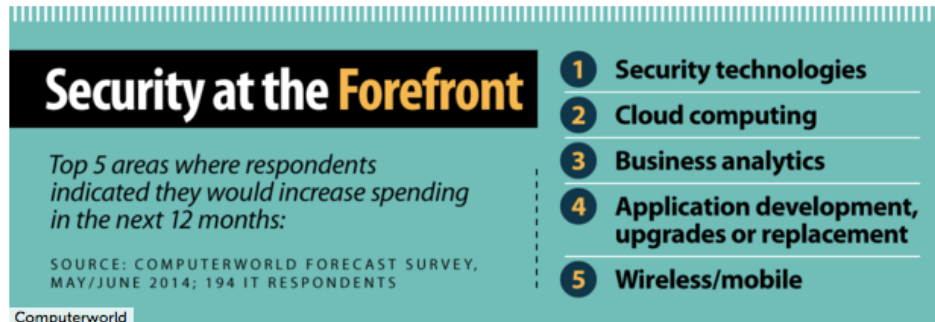


Why should you care about

INFORMATION RETRIEVAL?



Here's a look at what IT leaders say is on tap for 2015.



BIG DATA

Big Data: The Predictions For 2015
Bernard Marr

The 2 Most Powerful Words A Manager Can Use
Bernard Marr


Why You Need Emotional Intelligence To Succeed
Dr. Travis Bradberry

How To Avoid Big Company Disease How Toyota Got Their
Colin Baird


Give Yourself Space to Think
Rita J. King

My Best Job Search Tip: How To Create An 'Interview Bucket'
J.T. O'Donnell

3 Scripts that Keep Women Stuck at Work
Savitra Ford Walston



BIG DATA

Bernard Marr  **fluencer**
Best-Selling Author, Keynote Speaker and
Consultant in Strategy, Performance Management,
[Follow](#)

Companies and organizations will struggle to find data talent

There are expected to be 4.4 million people employed worldwide in positions directly involved with big data analysis by next year. But this won't be enough. By next year, 70% of US businesses will either have a data strategy in place or will be planning one for the near future, according to one survey. The number of colleges offering courses related to big data analysis continues to grow rapidly, but there will continue to be a shortage of workers trained in the necessary skills for the foreseeable future.

Textual analysis will become more widely used

Increasingly, much of the data we are storing for analysis is in an unstructured form. Textual analysis has become increasingly sophisticated in the last few years and that trend will continue. Computers will become more proficient at "reading" a piece of text (or voice converted to text) and spotting themes and sentiments – meaning it can be classified and analyzed in the same way as structured data.



Software Development Manager-Data Mining/Information Retrieval/Machine Learning

Amazon - Seattle, WA, US

Posted 23 hours ago

[Apply on company website](#)

Save



Other Details

About this job

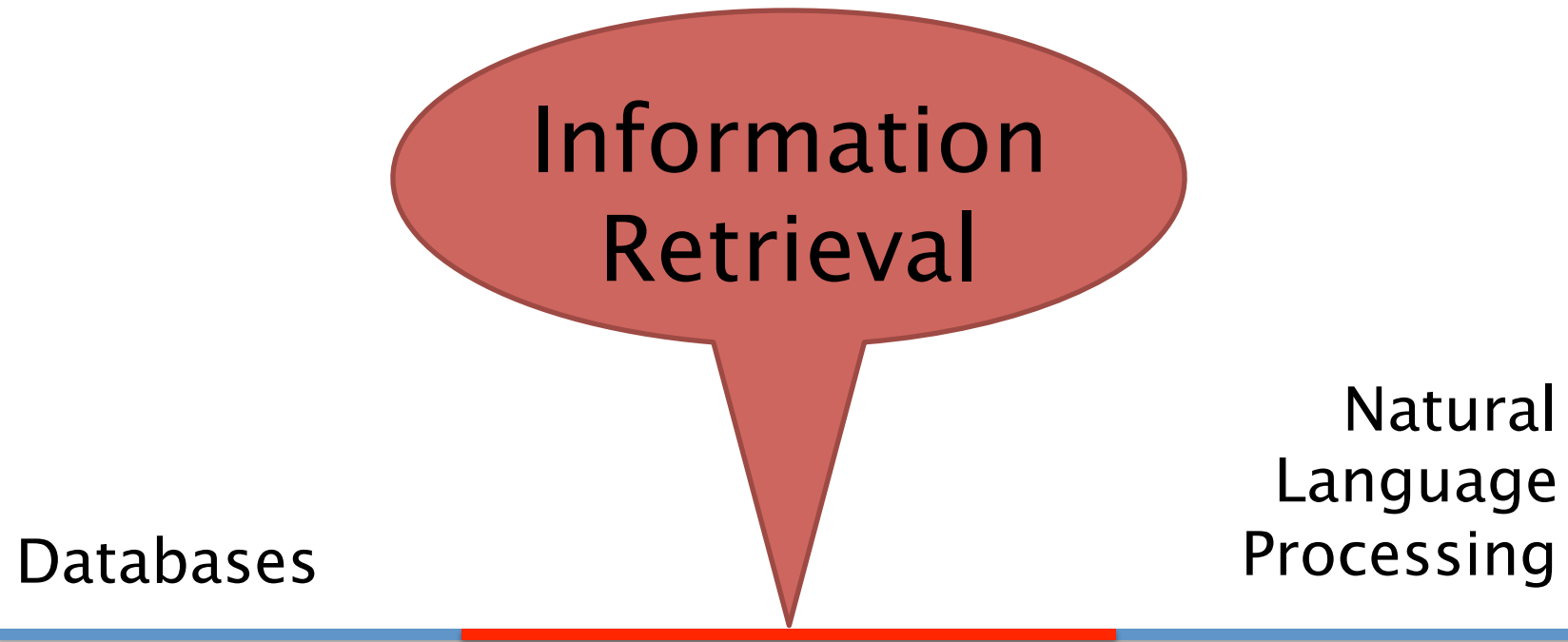


Job description

Listing Info

What do you like most about shopping at Amazon? Huge product selection? Amazon's catalog has billions of items. Despite its mind-boggling size and variety, it remains the best in the business. That's no accident! We use state of the art machine learning techniques and cloud computing technologies to keep it that way. A vast and high quality product catalog is a key strategic asset for Amazon that sets us apart from our competitors. Come help us make the world's best product catalog even better and influence the way millions of customers shop.

Right at the center of attention



How is it structured?





And now for the

COURSE ORGANIZATION

Lecturer

Min-Yen KAN

(as in $\min(x)$)

kanmy@comp.nus.edu.sg

<http://www.comp.nus.edu.sg/~kanmy>

Office: AS6 05-12 / COM1 02-19F
6516-1885

Hours: 13:00-14:00 Fri,
or by appointment

Hobbies:

Taking care of two little ones
Sleeping (-:



Teaching Assistants



Xu Chencan



chencan@comp.nus.edu.sg
xuchencan.wordpress.com

Office: AS6 #04-13
(Media Research Lab 2)
Hours: Thursday 2-5, AS6 4th
floor staff pantry and by
appointment
Hobbies: Writing poems,
Yoga, Regimen

Ashish Dandekar



ashishdandekar@u.nus.edu
www.facebook.com/profile.php?id=100001280016393

Office: AS6 #04-13
(Database Lab 1)
Hours: TBA and by appointment

Hobbies: Cooking, Travelling,
Reading



Course web sites

<http://www.comp.nus.edu.sg/~cs3245>

- Homework
- Other supplementary content

When using a search engine to find our course materials, make sure you find our site for 2014/15 Sem II.

<http://ivle.nus.edu.sg/>

- Lecture notes (via workbin)
- Discussion forum (participation counts!)
 - Any questions related to the course should be raised on this forum
 - Emails to me are considered public unless otherwise specified
- Announcements
- Homework submissions (via workbin)

Freedom of information rule



- Collaboration is acceptable and encouraged.
- To assure that all collaboration is on the level, **you must always fill in the name(s) of your collaborators on your assignment.**
- You will be assessed for the parts for which you claim is your own contribution.



Facebook rule

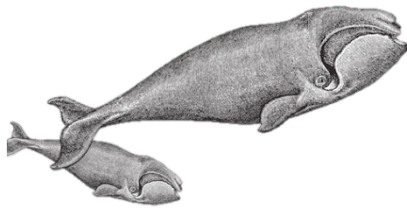
- You are free to meet with fellow students(s) and discuss assignments with them.
- Writing on a board or shared piece of paper is acceptable during the meeting; however, you **may not take any written (electronic or otherwise) record away from the meeting.**
- After the meeting, do something else for at least a half-hour (Facebook, or doing an assignment for a different class), before working on the assignment.
 - This will assure that you are able to reconstruct what you learned from the meeting, **by yourself, using your own brain.**
 - You will be asked to certify that you meet this requirement per assignment.

Python



- We'll be using the Python programming language for our entire class homework assignments.
- You are expected to learn Python and use it for the assignments.
- Don't worry, Python is an easy language to transition to for most programmers and it is very useful for string manipulation (a key characteristic for IR).

NLTK



- We will also be using the Natural Language Tool Kit (NLTK) for helping us deal with some parts of the course.
- It's coded in Python (surprise!)
 - Yes, we know that this is a course on **information retrieval** and not **natural language processing**, but the two are forever intertwined.
- Please install both Python 2.7.x or 3.2+ and NLTK on your own personal computer.
 - I will try to get a Programming Lab installed with the appropriate Python.